

## IS 2K-CONJECTURE VALID FOR FINITE VOLUME METHODS?

WAIXIANG CAO <sup>\*</sup>, ZHIMIN ZHANG <sup>†</sup>, AND QINGSONG ZOU <sup>‡</sup>

**Abstract.** This paper is concerned with superconvergence properties of a class of finite volume methods of arbitrary order over rectangular meshes. Our main result is to prove *2k-conjecture*: at each vertex of the underlying rectangular mesh, the bi- $k$  degree finite volume solution approximates the exact solution with an order  $O(h^{2k})$ , where  $h$  is the mesh size. As byproducts, superconvergence properties for finite volume discretization errors at Lobatto and Gauss points are also obtained. All theoretical findings are confirmed by numerical experiments.

**1. Introduction.** As a popular numerical method for partial differential equations (PDEs), the finite volume method (FVM) has a wide range of applications and attracts intensive theoretical studies, see, e.g., [3, 4, 6, 7, 14, 15, 17, 18, 19, 21, 23, 24, 26, 29, 33] for an incomplete list of publications. However, most theoretical studies in the literature have been focused on linear or quadratic schemes. Recently, arbitrary order FV schemes have been constructed and analyzed for elliptic problems in [8] and [30]. The basic idea of in [8, 30] to design a FV scheme of any order  $k$  is to choose standard finite element space as the trial space and construct control volumes with Gauss points in the primal partition. These FV schemes are shown to be convergent with optimal rates under both energy and  $L^2$  norms.

In 1973 Douglas-Dupont proved that the  $k$ th order  $C^0$  finite element method (FEM) to the two-point boundary value problem converges with rate  $h^{2k}$  at nodal points. Since then, it has been conjectured (based on many numerical evidences) that the same is true for bi- $k$  finite element approximation under rectangular meshes for the Poisson equation. This conjecture was settled (see [12]) recently after almost 40 years. Our earlier study reveals that a class of finite volume methods of arbitrary degree have similar (and even better in some special cases) superconvergence property as counterpart finite element methods in the one dimensional setting [8, 9]. It is natural to ask whether the *2k-conjecture* is valid for finite volume methods? In this work, we will provide a confirmatory answer to this question. To be more precise, we shall investigate superconvergence properties of any order FV schemes studied in [30]. In particular, we show that the underlying FVM has all superconvergence properties of the counterpart FEM.

We begin with a model problem:

$$-\Delta u = f \text{ in } \Omega, \text{ and } u = 0, \text{ on } \partial\Omega, \quad (1.1)$$

where  $\Omega = [a, b] \times [c, d]$  and  $f$  is a real-valued function defined on  $\Omega$ .

Techniques used in [8, 9] are very difficult to be applied to FV schemes in the two dimensional setting. Inspired by a recent work [12] for the finite element method, our approach here is to construct a suitable function to correct the error between the exact solution  $u$  and its interpolation  $u_I$ . Due to different nature of the finite

<sup>\*</sup> Beijing Computational Science Research Center, Beijing, 100084, China.

<sup>†</sup> Beijing Computational Science Research Center, Beijing, 100084, China. Department of Mathematics, Wayne State University, Detroit, MI 48202, USA. This author was supported in part by the US National Science Foundation through grant DMS-1115530.

<sup>‡</sup> College of Mathematics and Computational Science and Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou, 510275, P. R. China. This author is supported in part by the National Natural Science Foundation of China under the grant 11171359 and in part by the Fundamental Research Funds for the Central Universities of China.

volume method, the construction here is different from that of for the FEM, some novel design has to be made to serve our purpose. In particular, we construct our correction function by designing some special operators, instead of a complicated iterative procedure used in the FEM case (see Section 3). In addition, using a special mapping from the trial space to test space ([30]), the FV bilinear form can be regarded as a Gauss quadrature of its corresponding FE bilinear form. Then by taking special cares to the residual term of the Gauss quadrature, we show that our correction function also has desired properties. Once the correction function is constructed, superconvergence properties at some special points can be obtained with standard arguments. Our main results can be summarized as the following.

We first establish superconvergence at nodes : the bi- $k$  degree FV solution  $u_h$  superconverges to  $u$  with order  $2k$  at any nodal point  $P$ , i.e.,

$$(u - u_h)(P) = O(h^{2k}), \quad (\text{comparing with optimal global rate } O(h^{k+1})) \quad (1.2)$$

which is termed by Zhou and Lin ([31]) as  $2k$ -conjecture in the finite element regime, see also, e.g., [5, 27], for the literature along this line.

Our superconvergence results also include

$$(u - u_h)(L) = O(h^{k+2}), \quad (\text{comparing with } \|u - u_h\|_0 = O(h^{k+1})) \quad (1.3)$$

where  $L$  is an interior Lobatto point; and

$$\nabla(u - u_h)(G) = O(h^{k+1}), \quad (\text{comparing with } \|u - u_h\|_1 = O(h^k)) \quad (1.4)$$

where  $G$  is a Gauss point. As the reader may recall, these rates are the same as the counterpart FEM.

The rest of the paper is organized as follows. In Section 2, we present our FV scheme for (1.1) and discuss the relationship between FV and FE bilinear forms. Section 3 is the most technical part, where we construct a correction function and study its properties. In Section 4, we prove our main results (1.2) – (1.4). Finally, we provide some carefully designed numerical examples to support our theoretical findings in Section 5.

Throughout this paper, we adopt standard notations for Sobolev spaces such as  $W^{m,p}(D)$  on sub-domain  $D \subset \Omega$  equipped with the norm  $\|\cdot\|_{m,p,D}$  and semi-norm  $|\cdot|_{m,p,D}$ . When  $D = \Omega$ , we omit the index  $D$ ; and if  $p = 2$ , we set  $W^{m,p}(D) = H^m(D)$ ,  $\|\cdot\|_{m,p,D} = \|\cdot\|_{m,D}$ , and  $|\cdot|_{m,p,D} = |\cdot|_{m,D}$ . Notation “ $A \lesssim B$ ” implies that  $A$  can be bounded by  $B$  multiplied by a constant independent of the mesh size  $h$ . “ $A \sim B$ ” stands for “ $A \lesssim B$ ” and “ $B \lesssim A$ ”.

To end this introduction, we would like to emphasize that this work is a theoretical investigation. Our intention here is not to provide a practical method or anything like, rather, we settle a conjecture in convergence rate to the best possible case under very limited special situation.

Comparing with rich literature on superconvergence of the FEM (see, e.g., [2, 5, 10, 11, 20, 27, 25, 28, 32]), the superconvergence study for the FVM is still in its infancy, especially for high order schemes.

**2. Finite volume schemes of arbitrary order.** In this section, we first recall finite volume schemes introduced in [30], then we discuss briefly the relationship between the FV and its corresponding FE bilinear forms.

Let  $\mathcal{T}_h$  be a rectangular partition of  $\Omega$ , where  $h$  is the maximum length of all edges. For any  $\tau \in \mathcal{T}_h$ , we denote by  $h_\tau^x, h_\tau^y$  the lengths of  $x$ - and  $y$ - directional edges

of  $\tau$ , respectively. We assume that the mesh  $\mathcal{T}_h$  is *quasi-uniform* in the sense that there exist constants  $c_1, c_2 > 0$  such that

$$h \leq c_1 h_\tau^x, \quad h \leq c_2 h_\tau^y, \quad \forall \tau \in \mathcal{T}_h.$$

We denote by  $\mathcal{E}_h$  and  $\mathcal{N}_h$  the set of edges and vertices of  $\mathcal{T}_h$ , respectively.

We construct control volumes using Gauss points described below. Define reference element  $\hat{\tau} = [-1, 1] \times [-1, 1]$ , and  $\mathbb{Z}_r = \{1, 2, \dots, r\}$ ,  $\mathbb{Z}_r^0 = \{0, 1, \dots, r\}$  for all positive integer  $r$ . Let  $G_j, j \in \mathbb{Z}_k$  be Gauss points of degree  $k$  (zeros of the Legendre polynomial  $P_k$ ) in  $[-1, 1]$ . Then  $g_{i,j}^\tau = (G_i, G_j), i, j \in \mathbb{Z}_k$  constitutes  $k^2$  Gauss points in  $\hat{\tau}$ . Given  $\tau \in \mathcal{T}_h$ , let  $F_\tau$  be the affine mapping from  $\hat{\tau}$  to  $\tau$ . Then Gauss points in  $\tau$  are :

$$\mathcal{G}_\tau = \{g_{i,j}^\tau : g_{i,j}^\tau = F_\tau(g_{i,j}^{\hat{\tau}}), \quad i, j \in \mathbb{Z}_k\}.$$

Similarly, let  $L_i, i \in \mathbb{Z}_k^0$  be Lobatto points of degree  $k+1$  on the interval  $[-1, 1]$ , i.e.,  $L_0 = -1, L_k = 1$  and  $L_i, i \in \mathbb{Z}_{k-1}$  are zeros of  $P'_k$ . Then

$$\mathcal{N}_\tau = \{l_{i,j}^\tau : l_{i,j}^\tau = F_\tau(L_i, L_j), \quad i, j \in \mathbb{Z}_k^0\}$$

constitutes  $(k+1)^2$  Lobatto points on  $\tau$ . We denote by

$$\mathcal{N}^g = \bigcup_{\tau \in \mathcal{T}_h} \mathcal{G}_\tau, \quad \mathcal{N}^l = \bigcup_{\tau \in \mathcal{T}_h} \mathcal{N}_\tau$$

the set of Gauss and Lobatto points on the whole domain, respectively; and  $\mathcal{N}_0^l$  the set of interior Lobatto points by excluding Lobatto points on the boundary  $\partial\Omega$ . For any  $P \in \mathcal{N}_0^l$ , the control volume surrounding  $P$  is the rectangle  $K_P^*$  formed by four segments connecting the four Gauss points in  $\mathcal{N}^g$  closest to  $P$ . Then

$$\mathcal{T}_h^* = \bigcup_{P \in \mathcal{N}^l} K_P^*$$

constitutes a dual partition of  $\mathcal{T}_h$ .

Next, we denote  $\mathbb{P}_k$  as the space of polynomials with degree no more than  $k$ ; and  $\psi_{K_P^*}$ , the characteristic function of  $K_P^*$ . Then the trial and test spaces are defined as

$$U_h = \{v \in C(\Omega) : v|_\tau \in \mathbb{P}_k(x) \times \mathbb{P}_k(y), \tau \in \mathcal{T}_h, v|_{\partial\Omega} = 0\}$$

and

$$V_h = \text{Span}\{\psi_{K_P^*} : P \in \mathcal{N}_0^l\},$$

respectively. We see that  $U_h$  is the bi- $k$  degree finite element space, and  $V_h$  is the piecewise constants space with respect to the partition  $\mathcal{T}_h^*$ . They both vanish on the boundary of  $\Omega$ .

The finite volume method for solving (1.1) is to find  $u_h \in U_h$  satisfying the following local conservative property

$$-\int_{\partial\tau^*} \frac{\partial u_h}{\partial \mathbf{n}} ds = \int_{\tau^*} f dx dy, \quad \forall \tau^* \in \mathcal{T}_h^*,$$

or equivalently,

$$a_h(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_h, \quad (2.1)$$

where the bilinear form is defined for all  $w \in H_0^1(\Omega)$ ,  $v_h \in V_h$  by

$$a_h(w, v_h) = - \sum_{E \in \mathcal{E}_{\mathcal{T}_h^*}} [v_h]_E \int_E \frac{\partial w}{\partial \mathbf{n}} ds. \quad (2.2)$$

Here  $\mathcal{E}_{\mathcal{T}_h^*}$  is the set of interior edges of the dual partition  $\mathcal{T}_h^*$ ,  $[v_h]_E = v_h|_{\tau_2} - v_h|_{\tau_1}$  denotes the jump of  $v_h$  across the common edge  $E = \tau_1 \cap \tau_2$  of two rectangles  $\tau_1, \tau_2 \in \mathcal{T}_h^*$ , and  $\mathbf{n}$  denotes the normal vector on  $E$  pointing from  $\tau_1$  to  $\tau_2$ .

The inf-sup condition and continuity of the bilinear form  $a_h(\cdot, \cdot)$  have been established in [30]. Moreover, we have the following convergence and superconvergence properties.

LEMMA 2.1. (cf.[30]) *Let  $u \in H_0^1(\Omega) \cap H^{k+2}(\Omega)$  be the solution of (1.1), and  $u_h$ , the solution of (2.1). Then,*

$$|u - u_h|_1 \lesssim h^k |u|_{k+1}, \quad |u_h - \tilde{u}_I|_1 \lesssim h^{k+1} |u|_{k+2}, \quad (2.3)$$

where  $\tilde{u}_I \in U_h$  is the function interpolating  $u$  at Lobatto points.

We next discuss the relationship between  $a_h(\cdot, \cdot)$  and the FE bilinear form  $a_e(\cdot, \cdot)$ , which is defined for all  $v, w \in H^1(\Omega)$  by

$$a_e(v, w) = \int_{\Omega} \nabla v \cdot \nabla w.$$

We begin with some necessary notations. Let  $A_j, j \in \mathbb{Z}_k$  denote the weights of the Gauss quadrature  $Q_k(F) = \sum_{j=1}^k A_j F(G_j)$  for computing the integral  $I(F) = \int_{-1}^1 F(x) dx$ . For all  $\tau \in \mathcal{T}_h$  and  $v_1, v_2 \in L^2(\tau)$ , we define

$$\langle v_1, v_2 \rangle_{\tau} = \sum_{i,j=1}^k A_{\tau,i}^x A_{\tau,j}^y (v_1 v_2)(g_{i,j}^{\tau}),$$

where

$$A_{\tau,j}^x = \frac{1}{2} h_{\tau}^x A_j, \quad A_{\tau,j}^y = \frac{1}{2} h_{\tau}^y A_j, \quad j \in \mathbb{Z}_k$$

are Gauss weights associated with  $\tau$ . Then we can define a discrete inner product on  $\Omega$  :

$$\langle v_1, v_2 \rangle = \sum_{\tau \in \mathcal{T}_h} \sum_{i,j=1}^k A_{\tau,i}^x A_{\tau,j}^y v_1(g_{i,j}^{\tau}) v_2(g_{i,j}^{\tau}).$$

Writing  $\partial_x = \frac{\partial}{\partial x}, \partial_y = \frac{\partial}{\partial y}$  for simplicity, we denote, for all  $w \in H_0^1(\Omega)$ ,

$$\partial_x^{-1} w(x, y) = \int_a^x w(x', y) dx', \quad \partial_y^{-1} w(x, y) = \int_c^y w(x, y') dy'.$$

A function  $v_h \in V_h$  can be represented as

$$v_h = \sum_{P \in \mathcal{N}_0^l} (v_h)_P \psi_{K_P^*} = \sum_{P \in \mathcal{N}^l} (v_h)_P \psi_{K_P^*},$$

where  $(v_h)_P$  is a constant on the control volume  $K_P^*$  for  $P \in \mathcal{N}^l$ . Here we use the fact  $(v_h)_P = 0, P \in \partial\Omega$ .

Furthermore, we denote the (double layer) jump of  $v_h$  at the Gauss point  $g_{i,j}^\tau, \forall \tau \in \mathcal{T}_h, i, j \in \mathbb{Z}_k$  as

$$[v_h]_{g_{i,j}^\tau} = (v_h)_{l_{i,j}^\tau} + (v_h)_{l_{i-1,j-1}^\tau} - (v_h)_{l_{i-1,j}^\tau} - (v_h)_{l_{i,j-1}^\tau}.$$

With above notations, it is straightforward to deduce from (2.2) that

$$a_h(w, v_h) = - \sum_{\tau \in \mathcal{T}_h} \sum_{i,j=1}^k (\partial_x^{-1} \partial_y w + \partial_y^{-1} \partial_x w) (g_{i,j}^\tau) [v_h]_{g_{i,j}^\tau}. \quad (2.4)$$

In [30], a linear mapping  $\Pi : U_h \rightarrow V_h$

$$\Pi v = v_h =: \sum_{P \in \mathcal{N}_0^l} (v_h)_P \psi_{K_P^*} \in V_h, \quad v \in U_h, \quad (2.5)$$

is defined by letting

$$[v_h]_{g_{i,j}^\tau} = A_{\tau,i}^x A_{\tau,j}^y \partial_{xy}^2 v(g_{i,j}^\tau), \quad \forall g_{i,j}^\tau \in \mathcal{N}^g. \quad (2.6)$$

Note that although the number of constraints in (2.6) (which equals to the cardinality of  $\mathcal{N}^g$ ) is different from the dimensionality of the test space (which equals to the cardinality of  $\mathcal{N}_0^l$ ), it has been rigourously shown in [30] that  $\Pi$  is well-defined.

With this mapping, we have

$$a_h(w, \Pi v) = - \langle \partial_x^{-1} \partial_y w, \partial_{x,y}^2 v \rangle - \langle \partial_y^{-1} \partial_x w, \partial_{x,y}^2 v \rangle.$$

Since by Green's formula,

$$a_e(w, v) = - \int_{\Omega} (\partial_x^{-1} \partial_y w + \partial_y^{-1} \partial_x w) \partial_{x,y}^2 v dx dy,$$

therefore, the finite volume bilinear form  $a_h(\cdot, \Pi \cdot)$  can be regarded as the Gauss quadrature of the Galerkin bilinear form  $a_e(\cdot, \cdot)$ . Note that similar point of view appeared in the analysis of linear FV schemes in [18].

**3. Correction function.** Superconvergence analysis at a special point can usually be reduced to estimating

$$a_h(u - u_I, \Pi v), \forall v \in U_h,$$

where  $u_I \in U_h$  is an interpolant of  $u$  which will be defined in (3.10). A straightforward analysis using the continuity of  $a_h(\cdot, \cdot)$  results in

$$|a_h(u - u_I, \Pi v)| \lesssim h^k,$$

due to the restriction of optimal error bound

$$|u - u_I|_1 \lesssim h^k.$$

Further analysis based on standard superconvergence argument may lead to

$$|a_h(u - u_I, \Pi v)| \lesssim h^{k+1},$$

an improvement by order one, but is still far from our need. To obtain desired superconvergence results, more delicate analysis is necessary. In this section, we shall construct a correction function  $w_h$  with following properties.

**PROPOSITION 3.1.** *Assume that  $u \in H^{\alpha+1}(\Omega)$ ,  $\alpha = k + 2$  (or  $2k$ ). Then there exists a function  $w_h \in U_h$  such that  $w_h = 0$  at all nodes and*

$$\|w_h\|_{\infty} \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1}. \quad (3.1)$$

Furthermore,

$$|a_h(u - u_I - w_h, \Pi v)| \lesssim h^{\alpha} \|u\|_{\alpha+1} \|v\|_1, \quad \forall v \in U_h. \quad (3.2)$$

In the rest of this section, we will first construct  $w_h$  and then verify that  $w_h$  satisfies Proposition 3.1.

**3.1. Construction.** In this subsection, we construct a suitable correction function  $w_h$  by introducing some special operators. Our device is much transparent and simpler than that in [12] for the finite element method, where a complex iterative procedure is used.

We begin with notations and preliminaries. Since  $\mathcal{T}_h$  is a partition of rectangles, there exist  $a = x_0 < x_1 < \dots < x_m = b$  and  $c = y_0 < y_1 < \dots < y_n = d$  such that

$$\mathcal{T}_h = \{\tau_{i,j} : \tau_{i,j} = [x_{i-1}, x_i] \times [y_{j-1}, y_j], i \in \mathbb{Z}_m, j \in \mathbb{Z}_n\}.$$

We denote by  $B_i^x = [x_{i-1}, x_i] \times [c, d]$ ,  $i \in \mathbb{Z}_m$ , the element-band along  $x$ -direction and  $B_j^y = [a, b] \times [y_{j-1}, y_j]$ ,  $j \in \mathbb{Z}_n$ , the element-band along  $y$ -direction, respectively. For any rectangle  $B \subset \Omega$ , we define

$$U_h(B) = \{v \in C(\Omega) : v|_B \in \mathbb{P}_k(x) \times \mathbb{P}_k(y), v|_{\partial B} = 0\}.$$

Note that when  $k = 1$ ,  $U_h(B) = \{0\}$ .

For all  $i \in \mathbb{Z}_m$ , let  $\mathcal{L}_{B_i^x} : H_0^1(\Omega) \rightarrow U_h(B_i^x)$  be the operator which maps  $w \in H_0^1(\Omega)$  to  $\mathcal{L}_{B_i^x}(w)$  defined by

$$a_h(\mathcal{L}_{B_i^x}(w), \Pi v) = -\langle \partial_y^{-1} \partial_x w, \partial_{x,y}^2 v \rangle_{B_i^x}, \quad \forall v \in U_h(B_i^x). \quad (3.3)$$

Note that on one hand, given  $w \in H_0^1(\Omega)$ ,

$$-\langle \partial_y^{-1} \partial_x w, \partial_{x,y}^2 v \rangle_{B_i^x}, \quad \forall v \in U_h(B_i^x)$$

is a bounded linear functional on  $U_h(B_i^x)$ . On the other hand, the coercivity and continuity of the bilinear form  $a_h(\cdot, \Pi \cdot)$  have been established in [30]. Then by the Lax-Milgram Lemma, (3.3) has a unique solution and thus the operator  $\mathcal{L}_{B_i^x}$  is well defined.

We define a global operator  $\mathcal{L}^x : H_0^1(\Omega) \rightarrow U_h$  by

$$\mathcal{L}^x(w)|_{B_i^x} := \mathcal{L}_{B_i^x}(w), \quad \forall i \in \mathbb{Z}_m.$$

Since  $\mathcal{L}_{B_i^x}(w) = 0$  on the boundary  $\partial B_i^x$ ,  $\mathcal{L}^x(w) = 0$  on all  $\partial B_i^x$ ,  $i \in \mathbb{Z}_m$ . Consequently,  $\mathcal{L}^x(w) = 0$  at all vertices.

By a slight modification, we can define another operator  $\tilde{\mathcal{L}}^x : H_0^1(\Omega) \rightarrow U_h$  by letting

$$\tilde{\mathcal{L}}^x(w)|_{B_i^x} := \tilde{\mathcal{L}}_{B_i^x}(w), \quad \forall i \in \mathbb{Z}_m,$$

where the local operator  $\tilde{\mathcal{L}}_{B_i^x} : H_0^1(\Omega) \rightarrow U_h(B_i^x)$  is defined by

$$a_h(\tilde{\mathcal{L}}_{B_i^x}(w), \Pi v) = -\langle \partial_x^{-1} \partial_y w, \partial_{x,y}^2 v \rangle_{B_i^x}, \quad \forall v \in U_h(B_i^x). \quad (3.4)$$

By the same token, we define  $\mathcal{L}_{B_j^y}$ ,  $\tilde{\mathcal{L}}_{B_j^y}$ ,  $\mathcal{L}^y$ , and  $\tilde{\mathcal{L}}^y$ .

Next we define some projectors. Let  $P_r, r \geq 0$  be the Legendre polynomial of degree  $r$  and denote by

$$\phi_0(t) = \frac{1-t}{2}, \quad \phi_1(t) = \frac{1+t}{2}, \quad \phi_{r+1}(t) = \int_{-1}^t P_r(s) ds, \quad r \geq 1,$$

the series of Lobatto polynomials on the interval  $[-1, 1]$ . With these Lobatto polynomials, we have the following expansion for all  $v \in H^1(\Omega)$  and  $(x, y) \in B_i^x, i \in \mathbb{Z}_m$  along  $x$ -direction

$$v(x, y) = \sum_{r=0}^{\infty} b_r(y) \phi_r(s),$$

where  $s = (2x - x_i - x_{i-1})/h_i^x \in [-1, 1]$ ,

$$b_0(y) = v(x_{i-1}, y), \quad b_1(y) = v(x_i, y),$$

and

$$b_r(y) = \frac{2r-1}{2} \int_{-1}^1 \partial_s v(x, y) \phi_r'(s) ds, \quad r \geq 2. \quad (3.5)$$

Next, we define a projector  $Q_p^x, p \geq 1$  along the  $x$ -direction. Given  $(x, y) \in \Omega$ , there exists an  $i \in \mathbb{Z}_m$  such that  $(x, y) \in B_i^x$ , we then define

$$(Q_p^x v)(x, y) = \sum_{r=0}^p b_r(y) \phi_r(s).$$

Obviously,  $Q_p^x, p \geq 1$  is a bounded operator and  $Q_p^x v = v$  for all  $v(\cdot, y) \in \mathbb{P}_p$ . Consequently, by the Bramble-Hilbert lemma, there holds for all  $(x, y) \in B_i^x$

$$|(v - Q_p^x v)(x, y)| \lesssim h^p \int_{x_{i-1}}^{x_i} |\partial_x^{p+1} v(x, y)| dx \quad (3.6)$$

and

$$|\partial_x(v - Q_p^x v)(x, y)| \lesssim h^{p-1} \int_{x_{i-1}}^{x_i} |\partial_x^{p+1} v(x, y)| dx. \quad (3.7)$$

These inequalities will be frequently used in our later analysis. Moreover, by the properties of Legendre and Lobatto polynomials,

$$\partial_x(v - Q_p^x v)(\cdot, y) \perp \mathbb{P}_{p-1}, \quad (v - Q_p^x v)(\cdot, y) \perp \mathbb{P}_{p-2}, \quad \forall y \in [c, d], \quad (3.8)$$

where  $\mathbb{P}_{-1} = \emptyset$ . Noticing that  $\phi_r(\pm 1) = 0, r \geq 2$ , we have

$$(Q_p^x v)(x_i, y) = v(x_i, y), \quad (Q_p^x v)(x_{i-1}, y) = v(x_{i-1}, y), \quad \forall y \in [c, d]. \quad (3.9)$$

The projector  $Q_p^y, p \geq 1$  along  $y$ -direction can be defined similarly. With (3.9) and counterpart properties in the  $y$ -direction, we define an interpolation

$$v_I = Q_k^x Q_k^y v \quad (3.10)$$

and the residuals

$$E^x v = v - Q_k^x v, \quad E^y v = v - Q_k^y v,$$

then we have

$$(v - v_I)(P) = 0, \forall P \in \mathcal{N}_h,$$

and

$$v - v_I = E^x v + E^y v - E^y E^x v. \quad (3.11)$$

We are now in a perfect position to construct our correction function  $w_h$ . Let

$$w_h = \mathcal{L}^x(E^x u) + \mathcal{L}^y(E^y u) + \tilde{\mathcal{L}}^x(E^x u) + \tilde{\mathcal{L}}^y(E^y u) - \mathcal{L}^y(E^y E^x u) - \tilde{\mathcal{L}}^y(E^y E^x u). \quad (3.12)$$

Obviously,  $w_h \in U_h$  and  $w_h(P) = 0$  for all  $P \in \mathcal{N}_h$ .

**3.2. Analysis.** In this subsection, we shall prove  $w_h$  defined by (3.12) satisfies all properties listed in Proposition 3.1. For simplicity, we assume in this subsection that

$$h = h_\tau^x = h_\tau^y, \quad \forall \tau \in \mathcal{T}_h.$$

Consider  $\mathcal{L}^x(E^x u)$ , the first term of  $w_h$ . For this purpose, we need to present (3.3) in its linear algebraic form. We begin with a presentation of a basis of  $U_h(B_i^x), i \in \mathbb{Z}_m$ . For all  $(x, y) \in B_i^x$  and  $0 \leq p, q \leq k$ , let

$$\Psi_{p,q}(x, y) = \phi_p(s) \phi_q(t), \quad (3.13)$$

where

$$s = (2x - x_i - x_{i-1})/h, \quad t = (2y - d - c)/(d - c).$$

Then the function system  $\{\Psi_{p,q}, 2 \leq p, q \leq k\}$  constitutes a basis of  $U_h(B_i^x)$ . Since  $\mathcal{L}_{B_i^x}(E^x u) \in U_h(B_i^x)$ , we have the representation

$$\mathcal{L}_{B_i^x}(E^x u) = \sum_{p,q=2}^k w_{p,q} \Psi_{p,q}.$$

Let

$$D = (d_{p,q})_{(k-1) \times (k-1)}, \quad K = (m_{p,q})_{(k-1) \times (k-1)},$$

where

$$d_{p,q} = \langle \phi'_p, \phi'_q \rangle_{[-1, -1]}, \quad m_{p,q} = -\langle \partial^{-1} \phi_p, \phi'_q \rangle_{[-1, -1]}, \quad 2 \leq p, q \leq k$$



with the discrete inner product defined by

$$\langle v_1, v_2 \rangle_{[-1, -1]} = \sum_{r=1}^k A_r v_1(G_r) v_2(G_r).$$

By[16](p98, (2.7.12)),

$$\langle v_1, v_2 \rangle_{[-1, -1]} = \int_{-1}^1 (v_1 v_2)(x) dx - c_k (v_1 v_2)^{(2k)}(\xi), \quad (3.14)$$

where  $c_k = \frac{2^{2k+1}(k!)^4}{(2k+1)[(2k)!]^3}$  and  $\xi \in (-1, 1)$ . Taking  $v = \Psi_{r,l}, r, l = 2, \dots, k$  in (3.3), we derive

$$\sum_{p,q=2}^k \left( (d-c)^2 d_{p,r} m_{q,l} + h^2 d_{q,l} m_{p,r} \right) w_{p,q} = f_{r,l}, \quad (3.15)$$

where

$$f_{r,l} = -(d-c)h \langle \partial_y^{-1} \partial_x E^x u, \partial_{x,y}^2 \Psi_{r,l} \rangle_{B_i^x}. \quad (3.16)$$

Denote the unknowns  $X = (X_2, \dots, X_k)^T$  and the right-hand side  $F = (F_2, \dots, F_k)^T$  with vectors

$$X_r = (w_{r,2}, \dots, w_{r,k})^T, \quad F_r = (f_{r,2}, \dots, f_{r,k})^T, \quad r = 2, \dots, k.$$

Then (3.15) can be rewritten as

$$\left( (d-c)^2 (D \otimes K) + h^2 (K \otimes D) \right) X = F, \quad (3.17)$$

where for two matrices  $B_1 = (b_{p,q}^1)_{k \times k}$  and  $B_2 = (b_{p,q}^2)_{k \times k}$ , the tensor product  $B_1 \otimes B_2$  is a matrix of  $k^2 \times k^2$  defined by

$$B_1 \otimes B_2 = (B_{p,q})_{k \times k}, \quad B_{p,q} = b_{p,q}^1 B_2, \quad \forall p, q \leq k.$$

With the linear system (3.17), the study of the properties of  $\mathcal{L}^x(E^x u)$  is reduced to the estimation of the vector  $F$  and the matrix  $A = (d-c)^2 (D \otimes K) + h^2 (K \otimes D)$ .

We first estimate the vector  $F$ .

LEMMA 3.2. *If  $u \in H^{\alpha+1}(\Omega)$ ,  $\alpha \geq k+1$ , then*

$$\|F_p\|_\infty \lesssim h^{\min(\alpha, 2k+2-p)} |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1, B_i^x}, \quad p = 2, \dots, k. \quad (3.18)$$

*Proof.* For all  $\tau \in B_i^x, i \in \mathbb{Z}_m$ , we denote Gauss points  $g_{\tau,l}^\tau = (g_{\tau,r}^x, g_{\tau,l}^y), r, l \in \mathbb{Z}_k$ . Let  $\Theta = \partial_y^{-1} \partial_x (Q_{\alpha-1}^x E^x u)$ . Note that for any fixed  $y$ ,  $\partial_{x,y}^2 \Psi_{p,q}(\cdot, y) \in \mathbb{P}_{k-1}$ , by the orthogonality (3.8) and the fact that  $\Theta = \partial_x (Q_{\alpha-1}^x E^x (\partial_y^{-1} u))$ , we have

$$\int_{x_{i-1}}^{x_i} \Theta \partial_{x,y}^2 \Psi_{p,q} dx = \int_{x_{i-1}}^{x_i} (\partial_x E^x (\partial_y^{-1} u)) \partial_{x,y}^2 \Psi_{p,q} dx = 0,$$

thus

$$\langle \Theta, \partial_{x,y}^2 \Psi_{p,q} \rangle_{B_i^x} = - \sum_{\tau \in B_i^x} \sum_{l=1}^k A_{\tau,l}^y e_{p,q}^\tau (g_{\tau,l}^y),$$

where

$$e_{p,q}^\tau(y) = \int_{x_{i-1}}^{x_i} \Theta \partial_{x,y}^2 \Psi_{p,q} dx - \sum_{r=1}^k A_{\tau,r}^x (\Theta \partial_{x,y}^2 \Psi_{p,q}) (g_{\tau,r}^x, y)$$

is the error of Gauss quadrature for calculating the integral of  $\Theta \partial_{x,y}^2 \Psi_{p,q}$  in  $[x_{i-1}, x_i]$ . By (3.14), there exists a point  $\xi_i \in (x_{i-1}, x_i)$  such that

$$e_{p,q}^\tau(y) = c_k \frac{h^{2k+1}}{2^{2k+1}} \partial_x^{2k} (\Theta \partial_{x,y}^2 \Psi_{p,q}) (\xi_i, y).$$

Note that

$$\partial_y \Psi_{p,q} = \phi'_q = O(1), \quad \partial_x^{(r)} \Psi_{p,q} = \left(\frac{2}{h}\right)^r \phi_p^{(r)} = O(h^{-r}), \quad \forall r \leq p$$

and

$$\|\partial_x^j \Theta\|_{\infty, B_i^x} \lesssim \|\partial_x^{j+1} E^x(\partial_y^{-1} u)\|_{\infty, B_i^x} \lesssim |u|_{\alpha-1, \infty, B_i^x}, \quad \forall j < \alpha - 1.$$

Then, by the Leibnitz formula for derivatives,

$$|e_{p,q}^\tau| \lesssim h^{2k+1-p} |u|_{\alpha-1, \infty, B_i^x}, \quad 2 \leq q \leq k,$$

which implies

$$|\langle \Theta, \partial_{x,y}^2 \Psi_{p,q} \rangle_{B_i^x}| \lesssim h^{2k+1-p} |u|_{\alpha-1, \infty, B_i^x}. \quad (3.19)$$

On the other hand, by the approximation property of  $Q_p^x, p \geq 1$ ,

$$|E^x u - Q_{\alpha-1}^x E^x u|_{1, \infty, B_i^x} \lesssim h^{\alpha-1} |E^x u|_{\alpha, \infty, B_i^x} \lesssim h^{\alpha-1} |u|_{\alpha, \infty, B_i^x}.$$

Consequently,

$$|\langle \partial_y^{-1} \partial_x (E^x u - Q_{\alpha-1}^x E^x u), \partial_{x,y}^2 \Psi_{p,q} \rangle_{B_i^x}| \lesssim h^{\alpha-1} |u|_{\alpha, \infty, B_i^x}. \quad (3.20)$$

Furthermore, by the definition (3.16),

$$-\frac{f_{p,q}}{h(d-c)} = \langle \partial_y^{-1} \partial_x (E^x u - Q_{\alpha-1}^x E^x u), \partial_{x,y}^2 \Psi_{p,q} \rangle_{B_i^x} + \langle \Theta, \partial_{x,y}^2 \Psi_{p,q} \rangle_{B_i^x}.$$

Substituting (3.19) and (3.20) into the above equation, we obtain

$$|f_{p,q}|_\infty \lesssim h^{\min(\alpha, 2k+2-p)} \|u\|_{\alpha, \infty, B_i^x}.$$

Now recall from standard regularity argument [1],

$$\|u\|_{\alpha, \infty, B_i^x} \lesssim |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1, B_i^x},$$

the desired estimate (3.18) follows.  $\square$

We next study properties of the matrix  $A = (d-c)^2(D \otimes K) + h^2(K \otimes D)$ . By the orthogonality of Legendre polynomials and the fact that  $k$ -point Gauss quadrature is exact for polynomials of degree  $2k-1$ , we have

$$d_{p,q} = (P_{p-1}, P_{q-1}) = 0, p \neq q, \quad d_{p,p} = \frac{2}{2p-1}, \quad p, q = 2, \dots, k.$$

In other words,  $D$  is a diagonal matrix. Similarly,

$$m_{p,q} = -(\partial^{-1}\phi_p, \phi'_q) = (\phi_p, \phi_q), \quad p, q \leq k, p+q \leq 2k-1. \quad (3.21)$$

By the quasi-orthogonal property of Lobatto polynomials,  $m_{p,q} \neq 0$  only when  $p-q = 0, \pm 2$ . Consequently,  $K$  is a five-diagonal matrix.

LEMMA 3.3. *The matrix  $K$  is symmetric and positive definite.*

*Proof.* Let  $K_1 = (m_{p,q}^1)_{(k-1) \times (k-1)}$  with  $m_{p,q}^1 = (\phi_p, \phi_q), p, q = 2, \dots, k$ . By (3.21),

$$m_{p,q}^1 = m_{p,q}, \quad \forall p, q \leq k, p+q \leq 2k-1.$$

We next study the relationship of  $m_{k,k}^1$  and  $m_{k,k}$ . Denoting

$$e_k = m_{k,k} - m_{k,k}^1,$$

we have from (3.14) and the Leibnitz formula for derivatives

$$e_k = c_k((\partial^{-1}\phi_k)\phi'_k)^{(2k)}(\xi) = c_k \binom{2k}{k-1} \|\phi_k\|_{k,\infty}^2 > 0.$$

Then

$$K = K_1 + K_2,$$

where  $K_2 = (m_{p,q}^2)_{(k-1) \times (k-1)}, p, q = 2, \dots, k$  with

$$m_{k,k}^2 = e_k > 0, \quad m_{p,q}^2 = 0, \text{ otherwise.}$$

Since  $K_1$  is symmetric and positive definite,  $K$  is also symmetric and positive definite.

□

Note that both  $D$  and  $K$  are symmetric and positive definite and independent of  $h$ , then both  $D \otimes K$  and  $D \otimes K$  are also positive definite. By the definition of  $A$ , we have

$$\det(A) = \det((d-c)^2(D \otimes K)) + O(h^2).$$

Therefore, when  $h$  is sufficiently small,  $\det A$  is positive and uniformly bounded from below. In other words, when  $h$  is sufficiently small

$$0 < \det(A)^{-1} \lesssim C, \quad (3.22)$$

where  $C$  is independent of  $h$ .

With the estimate for  $F$  and properties of  $A$ , we are now ready to estimate  $\mathcal{L}_{B_i^x}(E^x u)$ .

LEMMA 3.4. *Assume  $u \in H^{\alpha+1}(\Omega), \alpha = k+2$  (or  $2k$ ). Then for sufficiently small  $h$  and all  $i \in \mathbb{Z}_m$*

$$\|X_r\|_\infty \lesssim h^{k+2+\max(0, \alpha-k-r)} |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1, B_i^x}, \quad r = 2, \dots, k. \quad (3.23)$$

Consequently,

$$\|\mathcal{L}_{B_i^x}(E^x u)\|_\infty \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1, B_i^x}. \quad (3.24)$$

*Proof.* Note that

$$\|\mathcal{L}_{B_i^x}(E^x u)\|_\infty \lesssim \sum_{r=2}^k \|X_r\|_\infty,$$

then (3.24) follows from (3.23). We next show (3.23). When  $u \in H^{k+3}(\Omega)$ , By (3.18), (3.22) and the Cramer's rule, we have

$$\|X_r\|_\infty \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{k+3, B_i^x}, \quad r = 2, \dots, k.$$

Then (3.23) is valid for  $\alpha = k + 2$ . To prove (3.23) for the case  $\alpha = 2k$ , we rewrite  $A$  in its block matrix form  $A = (A_{r,l})_{(k-1) \times (k-1)}$ , where each

$$A_{r,l} = (d - c)^2 d_{r,l} K + h^2 m_{r,l} D, \quad r, l = 2, \dots, k$$

is a  $(k-1) \times (k-1)$  matrix. Let

$$A'_{r,l} = A_{r,l} h^{-|r-l|}$$

and

$$Y_r = X_r h^{r-2k-2} |\ln h|^{-\frac{1}{2}} \|u\|_{2k+1, B_i^x}^{-1}, \quad F'_r = F_r h^{r-2k-2} |\ln h|^{-\frac{1}{2}} \|u\|_{2k+1, B_i^x}^{-1}.$$

Then both  $A'_{r,l}$  and  $F'_r$  are independent of  $h$ . By (3.18), we have

$$\|F_r\|_\infty \lesssim h^{2k+2-r} |\ln h|^{\frac{1}{2}} \|u\|_{2k+1, B_i^x}.$$

Multiplying the  $r$ -th equation of (3.17) with the factor  $h^{r-2k-2} |\ln h|^{-\frac{1}{2}} \|u\|_{2k+1, B_i^x}^{-1}$ , we have for all  $r = 2, \dots, k$

$$h^4 A'_{r,r-2} Y_{r-2} + A'_{r,r} Y_r + A'_{r,r+2} Y_{r+2} = F'_r, \quad (3.25)$$

where we use the notations  $A_{2,0} = A_{3,1} = A_{k-1,k+1} = A_{k,k+2} = 0$ . Let  $B = (B_{r,l})_{(k-1) \times (k-1)}$  with

$$B_{r,l} = A'_{r,l}, \quad r \leq l, \quad B_{r,l} = h^4 A'_{r,l}, \quad \text{otherwise.}$$

Then (3.25) can be written as a linear system  $BY = F'$ . A direct calculation yields

$$\det(B) = \prod_{r=2}^k \det(A'_{r,r}) + O(h^4),$$

which means that  $B$  is uniformly bounded from below. By Cramer's rule, each entry of  $Y$  is bounded independent of  $h$ . In other words,  $\|Y_r\|_\infty \lesssim 1$ . Consequently,

$$\|X_r\|_\infty \lesssim h^{2k+2-r} |\ln h|^{\frac{1}{2}} \|u\|_{2k+1, B_i^x}, \quad r = 2, \dots, k.$$

This finishes our proof.  $\square$

To prove Proposition 3.1, we still need to study the residual

$$(R_i^x(w), v) = -\langle \partial_y^{-1} \partial_x w, \partial_{x,y}^2 v \rangle_{B_i^x} - a_h(\mathcal{L}_{B_i^x}(w), \Pi v), \quad w \in H_0^1(\Omega)$$

for a general function  $v \in U_h$ . Note that when  $v \in U_h(B_i^x)$ , we have  $(R_i^x(w), v) = 0$ .

LEMMA 3.5. Assume that  $u \in H^{\alpha+1}(\Omega)$ ,  $\alpha = k + 2$  (or  $2k$ ). Then for a general function  $v \in U_h$ ,

$$|(R_i^x(E^x u), v)| \lesssim h^\alpha \|u\|_{\alpha+1, B_i^x} \|v\|_{1, B_i^x}, \quad \forall i \in \mathbb{Z}_m. \quad (3.26)$$

*Proof.* Note that for all  $B_i^x \subset \Omega$ ,  $i \in \mathbb{Z}_m$ ,

$$\phi_0(s) = \frac{x_i - x}{h}, \quad \phi_1(s) = \frac{x - x_{i-1}}{h} \notin U_h(B_i^x),$$

where  $s = (2x - x_i - x_{i-1})/h \in [-1, 1]$ . Then a general function  $v \in U_h$  has the decomposition

$$v(x, y) = v_h(x, y) + \tilde{v}(x, y), \quad \forall (x, y) \in B_i^x,$$

where  $v_h \in U_h(B_i^x)$  and  $\tilde{v}(x, y) = v(x_{i-1}, y)\phi_0(s) + v(x_i, y)\phi_1(s)$ . By (3.3),

$$(R_i^x(E^x u), v) = -\langle \partial_y^{-1} \partial_x E^x u, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x} - a_h(\mathcal{L}_{B_i^x}(E^x u), \Pi \tilde{v}) = -J_1 - J_2.$$

We next estimate  $J_1$  and  $J_2$  separately. Let  $\Phi = \partial_y^{-1} \partial_x E^x u$ . By (3.6)-(3.7) and the fact that  $\partial_y^{k+1} \Phi = \partial_x E^x (\partial_y^k u)$ , we have for all  $(x, y) \in \tau_{i,j}$ ,  $(i, j) \in \mathbb{Z}_m \times \mathbb{Z}_n$

$$\begin{aligned} |(\Phi - Q_k^y \Phi)(x, y)| &\lesssim h^k \int_{y_{j-1}}^{y_j} |\partial_x E^x (\partial_y^k u)(x, y)| dy \\ &\lesssim h^{\alpha-1} \int_{y_{j-1}}^{y_j} \int_{x_{i-1}}^{x_i} |\partial_x^{\alpha+1-k} E^x (\partial_y^k u)(x, y)| dx dy \lesssim h^\alpha |u|_{\alpha+1, \tau_{i,j}}. \end{aligned}$$

Then by the Cauchy-Schwartz inequality, we derive

$$\begin{aligned} |\langle \Phi - Q_k^y \Phi, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x}| &\lesssim \langle \Phi - Q_k^y \Phi, \Phi - Q_k^y \Phi \rangle_{B_i^x}^{\frac{1}{2}} \langle \partial_{x,y}^2 \tilde{v}, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x}^{\frac{1}{2}} \\ &\lesssim h^\alpha |u|_{\alpha+1, B_i^x} \|\partial_x \tilde{v}\|_{0, B_i^x}. \end{aligned}$$

Here in the last step, we have used the inverse inequality

$$\|\partial_{x,y}^2 \tilde{v}\|_{0, B_i^x} \lesssim h^{-1} \|\partial_x \tilde{v}\|_{0, B_i^x}.$$

Note that  $(Q_k^y \Phi) \partial_{x,y}^2 \tilde{v}(x, \cdot) \in \mathbb{P}_{2k-1}$ , by Gauss quadrature and integrating by part, we obtain

$$\begin{aligned} \langle Q_k^y \Phi, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x} &= - \sum_{\tau \in B_i^x} \sum_{l=1}^k A_{\tau,l}^x \int_c^d (\partial_y Q_k^y \Phi) \partial_x \tilde{v}(g_{\tau,l}^x, y) dy \\ &= - \sum_{\tau \in B_i^x} \sum_{l=1}^k A_{\tau,l}^x \int_c^d \partial_x E^x (\partial_y Q_k^y \partial_y^{-1} u) \partial_x \tilde{v}(g_{\tau,l}^x, y) dy. \end{aligned}$$

Let  $\Upsilon = \partial_y Q_k^y \partial_y^{-1} u$ . Since  $\tilde{v}$  is linear with respect to  $x$ , we have

$$\sum_{\tau \in B_i^x} \sum_{l=1}^k A_{\tau,l}^x \int_c^d (\partial_x Q_\alpha^x E^x \Upsilon) \partial_x \tilde{v}(g_{\tau,l}^x, y) dy = \int_{B_i^x} (\partial_x Q_\alpha^x E^x \Upsilon) \partial_x \tilde{v} dx dy = 0.$$

Consequently,

$$\begin{aligned} |\langle Q_k^y \Phi, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x}| &= \sum_{\tau \in B_i^x} \sum_{l=1}^k A_{\tau,l}^x \int_c^d |((\partial_x E^x \Upsilon - \partial_x Q_\alpha^x E^x \Upsilon) \partial_x \tilde{v})(g_{\tau,l}^x, y)| dy \\ &\lesssim h^\alpha \|\partial_x^{\alpha+1} E^x \Upsilon\|_{0, B_i^x} \|\partial_x \tilde{v}\|_{0, B_i^x} \lesssim h^\alpha |u|_{\alpha+1, B_i^x} \|\partial_x \tilde{v}\|_{0, B_i^x}. \end{aligned}$$

Note that

$$\partial_x \tilde{v} = \frac{v(x_i, y) - v(x_{i-1}, y)}{h} = h^{-1} \int_{x_{i-1}}^{x_i} \partial_x v(x, y) dx,$$

we have

$$\|\partial_x \tilde{v}\|_{0, B_i^x} \lesssim \|v\|_{1, B_i^x}.$$

Then

$$\begin{aligned} |J_1| &= |\langle \Phi - Q_k^y \Phi, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x} + \langle Q_k^y \Phi, \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x}| \\ &\lesssim h^\alpha |u|_{\alpha+1, B_i^x} \|v\|_{1, B_i^x}. \end{aligned}$$

As for  $J_2$ , recall the bilinear form  $a_h(\cdot, \Pi \cdot)$ , and we have

$$J_2 = -\langle \partial_y^{-1} \partial_x \mathcal{L}_{B_i^x}(E^x u), \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x} - \langle \partial_x^{-1} \partial_y \mathcal{L}_{B_i^x}(E^x u), \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x}$$

Note that

$$\int_{x_{i-1}}^{x_i} (\partial_{x,y}^2 \tilde{v}) \partial_y^{-1} \partial_x \mathcal{L}_{B_i^x}(E^x u) dx = 0,$$

then

$$\langle \partial_y^{-1} \partial_x \mathcal{L}_{B_i^x}(E^x u), \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x} = 0.$$

Therefore,

$$J_2 = -\langle \partial_x^{-1} \partial_y \mathcal{L}_{B_i^x}(E^x u), \partial_{x,y}^2 \tilde{v} \rangle_{B_i^x} = -\int_{B_i^x} \frac{\partial^2 \tilde{v}}{\partial x \partial y} \partial_x^{-1} \partial_y \mathcal{L}_{B_i^x}(E^x u) dx dy.$$

Since

$$\left( \partial_x^{-1} \partial_y \mathcal{L}_{B_i^x}(E^x u) \right)(x_i) = \left( \partial_x^{-1} \partial_y \mathcal{L}_{B_i^x}(E^x u) \right)(x_{i-1}) = 0, \quad \tilde{v}(x, c) = \tilde{v}(x, d) = 0,$$

integrating by part, we obtain

$$\begin{aligned} J_2 &= -\int_{B_i^x} \tilde{v} \left( \partial_y^2 \mathcal{L}_{B_i^x}(E^x u) \right) dx dy \\ &= -\sum_{p,q=2}^k w_{p,q} \int_{B_i^x} \left( v(x_{i-1}, y) \phi_0(s) + v(x_i, y) \phi_1(s) \right) \partial_y^2 \Psi_{p,q} dx dy. \end{aligned}$$

Note that  $\Psi_{p,q}(\cdot, y) \perp \mathbb{P}_1, p > 3$ , then only  $p = 2, 3$  in the above equation remain. For any  $q = 2, \dots, k$ , a direct calculation yields

$$\begin{aligned} \left| \int_{B_i^x} \left( v(x_{i-1}, y) \phi_0(s) + v(x_i, y) \phi_1(s) \right) \partial_y^2 \Psi_{2,q} dx dy \right| &\lesssim h \int_c^d |v(x_{i-1}, y) + v(x_i, y)| dy \\ &\lesssim \int_{B_i^x} |v(x, y)| dy. \end{aligned}$$

Here in the last step, we have used the inverse inequality

$$|v(\xi_i, y)| \lesssim h^{-1} \int_{x_{i-1}}^{x_i} |v(x, y)| dx, \quad \forall \xi_i \in [x_{i-1}, x_i], v \in U_h.$$

By the same argument, we derive

$$\begin{aligned} \left| \int_{B_i^x} \left( v(x_{i-1}, y) \phi_0(s) + v(x_i, y) \phi_1(s) \right) \partial_y^2 \Psi_{3,q} dx dy \right| &\lesssim h \int_c^d |v(x_{i-1}, y) - v(x_i, y)| dy \\ &\lesssim h \int_{B_i^x} |\partial_x v(x, y)| dy. \end{aligned}$$

Substituting the above two inequalities into the formula of  $J_2$ , we have

$$\begin{aligned} |J_2| &\lesssim (\|X_2\|_\infty + h\|X_3\|_\infty) \|v\|_{1,1,B_i^x} \\ &\lesssim h^{\alpha+\frac{1}{2}} |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1,B_i^x} \|v\|_{1,B_i^x}. \end{aligned}$$

Then the desired result follows by combining  $J_1$  with  $J_2$ .  $\square$

Similarly, by denoting the residual for all  $j \in \mathbb{Z}_n$

$$(R_j^y(w), v) = -\langle \partial_y^{-1} \partial_x w, \partial_{x,y}^2 v \rangle_{B_j^y} - a_h(\mathcal{L}_{B_j^y}(w), \Pi v), \quad w \in H_0^1(\Omega), v \in U_h,$$

we have

$$|(R_j^y(E^y u), v)| \lesssim h^\alpha \|u\|_{\alpha+1,B_j^y} \|v\|_{1,B_j^y}, \quad (3.27)$$

and

$$|(R_j^y(E^y E^x u), v)| \lesssim h^\alpha \|u\|_{\alpha+1,B_j^y} \|v\|_{1,B_j^y}. \quad (3.28)$$

With all the above preparations, we are ready to prove Proposition 3.1.

*Proof of Proposition 3.1.* As a direct consequence of (3.24), we have

$$\|\mathcal{L}^x(E^x u)\|_\infty \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{\alpha+1}.$$

Similar results hold true for  $\mathcal{L}^y(E^y u)$ ,  $\tilde{\mathcal{L}}^x(E^x u)$ ,  $\tilde{\mathcal{L}}^y(E^y u)$  and  $\mathcal{L}^y(E^y E^x u)$ ,  $\tilde{\mathcal{L}}^y(E^y E^x u)$  by the same arguments. Then (3.1) follows.

Now we turn to prove (3.2). Let  $R = u - u_I$ . By the orthogonal property, we have for all  $v \in U_h$

$$\begin{aligned} a_h(u_h - u_I, \Pi v) &= a_h(u - u_I, \Pi v) \\ &= -\langle \partial_y^{-1} \partial_x R, \partial_{x,y}^2 v \rangle - \langle \partial_x^{-1} \partial_y R, \partial_{x,y}^2 v \rangle = I_1 + I_2. \end{aligned}$$

From the decomposition (3.11), we have

$$I_1 = -\langle \partial_y^{-1} \partial_x E^x u, \partial_{x,y}^2 v \rangle - \langle \partial_y^{-1} \partial_x E^y u, \partial_{x,y}^2 v \rangle + \langle \partial_y^{-1} \partial_x (E^y E^x u), \partial_{x,y}^2 v \rangle.$$

Let  $w_h = w_1 + w_2$  with

$$w_1 = \mathcal{L}^x(E^x u) + \mathcal{L}^y(E^y u) - \mathcal{L}^y(E^y E^x u),$$

and

$$w_2 = \tilde{\mathcal{L}}^x(E^x u) + \tilde{\mathcal{L}}^y(E^y u) - \tilde{\mathcal{L}}^y(E^y E^x u).$$

By (3.26)-(3.28), we derive

$$\begin{aligned} |I_1 - a_h(w_1, \Pi v)| &= \sum_{B_i^x} |(R_i^x(E^x u), v)| + \sum_{B_j^y} |(R_j^y(E^y u), v)| + |(R_j^y(E^y E^x u), v)| \\ &\lesssim h^\alpha \|u\|_{\alpha+1} \|v\|_1. \end{aligned}$$

By the same arguments, we have

$$|I_2 - a_h(w_2, \Pi v)| \lesssim h^\alpha \|u\|_{\alpha+1} \|v\|_1.$$

Note that

$$a_h(u - u_I - w_h, \Pi v) = I_1 - a_h(w_1, \Pi v) + I_2 - a_h(w_2, \Pi v),$$

then (3.2) follows.  $\square$

**4. Superconvergence.** In this section, we shall study superconvergence properties of  $u_h$  at three kinds of special points : nodes, Gauss and Lobatto points.

Our first goal is to prove the *2k-conjecture*.

**THEOREM 4.1.** *Let  $u \in H^{2k+1}(\Omega)$  be the solution of (1.1), and  $u_h$  the solution of (2.1). Then,*

$$|(u - u_h)(P)| \lesssim h^{2k} |\ln h|^{\frac{1}{2}} \|u\|_{2k+1}, \forall P \in \mathcal{N}_h. \quad (4.1)$$

*Proof.* By [30], there hold

$$a_h(w, \Pi v) \lesssim \|w\|_1 \|v\|_1, \quad a_h(v, \Pi v) \gtrsim \|v\|_1^2, \quad \forall w, v \in U_h. \quad (4.2)$$

For any  $v \in U_h$  and  $Q \in \Omega$ , by the Lax-Milgram Lemma, there exists  $g_h \in U_h$  such that

$$a_h(v, \Pi g_h) = v(Q). \quad (4.3)$$

Choosing  $v = g_h$ , we have, from (4.2) and (4.3)

$$\|g_h\|_1^2 \leq |a_h(g_h, \Pi g_h)| = |g_h(Q)| \leq \|g_h\|_\infty.$$

Since (cf., [32], p.84, Theorem 2.8)

$$\|v\|_\infty \lesssim |\ln h|^{\frac{1}{2}} \|v\|_1, \quad \forall v \in U_h,$$

we have

$$\|g_h\|_1 \lesssim |\ln h|^{\frac{1}{2}}. \quad (4.4)$$

Letting  $v = u_h - u_I - w_h \in U_h$  in (4.3) and using (3.2) and (4.4), we obtain

$$|(u_h - u_I - w_h)(Q)| = |a_h(u - u_I - w_h, \Pi g_h)| \lesssim h^{2k} |\ln h|^{\frac{1}{2}} \|u\|_{2k+1}. \quad (4.5)$$



Noticing  $w_h = 0$  and  $u_I = u$  at all nodes  $P \in \mathcal{N}_h$ , the desired result (4.1) follows.  $\square$

We next discuss superconvergence of  $u_h$  at Gauss and Lobatto points.

**THEOREM 4.2.** *Let  $u \in H^{k+3}(\Omega)$  be the solution of (1.1), and  $u_h$  the solution of (2.1). Then,*

$$|(u - u_h)(P)| \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{k+3}, \quad \forall P \in \mathcal{N}^l, \quad (4.6)$$

and

$$|\nabla(u - u_h)(Q)| \lesssim h^{k+1} |\ln h|^{\frac{1}{2}} \|u\|_{k+3}, \quad \forall Q \in \mathcal{N}^g. \quad (4.7)$$

*Proof.* By (3.1)-(3.2) and (4.3), we have

$$\|u_I - u_h\|_\infty \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{k+3}.$$

By the inverse inequality,

$$|u_I - u_h|_{1,\infty} \lesssim h^{-1} \|u_I - u_h\|_\infty \lesssim h^{k+1} |\ln h|^{\frac{1}{2}} \|u\|_{k+3}.$$

On the other hand, by the definition of  $u_I$ , we have (see, e.g., [10, 32])

$$|(u - u_I)(P)| \lesssim h^{k+2} |u|_{k+2,\infty}, \quad \forall P \in \mathcal{N}^l,$$

and

$$|\nabla(u - u_I)(Q)| \lesssim h^{k+1} |u|_{k+2,\infty}, \quad \forall Q \in \mathcal{N}^g.$$

The desired statements (4.6)-(4.7) then follows.  $\square$

**REMARK 4.3.** *As a direct consequence of the above theorem, we have*

$$|u_h - u_I|_1 \lesssim |u_h - u_I|_{1,\infty} \lesssim h^{k+1} |\ln h|^{\frac{1}{2}} \|u\|_{k+3},$$

and

$$\|u_I - u_h\|_0 \lesssim \|u_I - u_h\|_\infty \lesssim h^{k+2} |\ln h|^{\frac{1}{2}} \|u\|_{k+3}.$$

*It was pointed out in [30] that the FV approximation  $u_h$  is super-close to the Lobatto interpolation function  $\tilde{u}_I$ . The above inequalities clearly indicate the same for the interpolation function  $u_I$ , i.e.,  $u_h$  is also super-close to  $u_I$  up to a logarithmic factor.*

**5. Numerical results.** In this section, we present numerical examples to support our theoretical findings in the previous section.

We consider (1.1) with  $\Omega = [0, 1] \times [0, 1]$  and the right-hand side

$$\begin{aligned} f(x, y) = & [(5\pi^2 - 4y^2 - 3) \sin(\pi x) \sin(2\pi y) - 8\pi y \sin(\pi x) \cos(2\pi y) \\ & - 2\pi \cos(\pi x) \sin(2\pi y)] e^{x-0.5+y^2}. \end{aligned}$$

The exact solution is then

$$u(x, y) = \sin(\pi x) \sin(2\pi y) e^{x-0.5+y^2}, \quad (x, y) \in \Omega.$$

We construct  $\mathcal{T}_h$  with  $h = 2^{-s}$ ,  $s = 1, 2, \dots, 8$ , by dividing  $\Omega$  into  $h^{-1} \times h^{-1}$  squares, and solve (1.1) by the FV scheme (2.1) with  $k = 3, 4$ . For each  $h$  and  $k$ , we

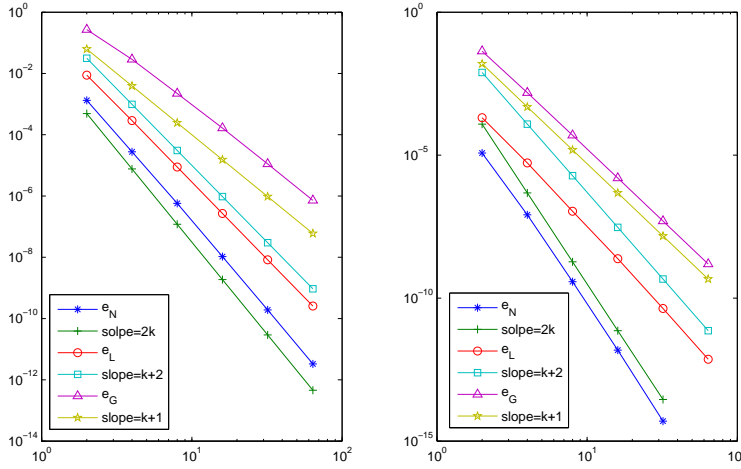
TABLE 5.1

	$k = 3$			$k = 4$		
N	$e_G$	$e_L$	$e_N$	$e_G$	$e_L$	$e_N$
2	2.699e-1	8.851e-3	1.327e-3	4.326e-2	2.044e-4	1.190e-5
4	2.897e-2	2.902e-4	2.761e-5	1.536e-3	5.354e-6	8.178e-8
8	2.224e-3	8.863e-6	5.743e-7	4.979e-5	1.092e-7	3.750e-10
16	1.660e-4	2.701e-7	1.056e-8	1.586e-6	2.397e-9	1.510e-12
32	1.117e-5	8.288e-9	1.919e-10	4.986e-8	4.340e-11	—
64	7.222e-7	2.567e-10	3.309e-12	1.564e-9	7.257e-13	—

measure maximum errors at nodes, Lobatto points, and Gauss points (for gradient only), respectively. They are defined by

$$e_N = \max_{P \in \mathcal{N}_h} |(u - u_h)(P)|, \quad e_L = \max_{P \in \mathcal{N}^l} |(u - u_h)(P)|, \quad e_G = \max_{Q \in \mathcal{N}^g} |\nabla(u - u_h)(Q)|.$$

Numerical data are demonstrated in Table 5.1, and corresponding error curves are depicted in Figure 5.1 with log-log scale. We observe a convergence slope  $k + 1$  for  $e_G$ ,  $k + 2$  for  $e_L$ , and  $2k$  for  $e_N$ , respectively. These results confirm our theoretical findings in Theorem 4.2 and Theorem 4.1: The derivative error is superconvergent at all Gauss points and the function value error is superconvergent at all Lobatto points. Moreover, the approximation error at nodes converges with a rate  $h^{2k}$ , the  $2k$ -conjecture for our finite volume approximation is verified.

FIG. 5.1. left:  $k = 3$ , right:  $k = 4$ .

## REFERENCES

- [1] R. A. Adams. *Sobolev spaces*, Academic Press, New York, 1975.
- [2] I. Babuška and T. Strouboulis and C. S. Upadhyay and S. K. Gangaraj. Computer-based proof of the existence of superconvergence points in the finite element method : superconvergence

- of the derivatives in finite element solutions of Laplace's, Poisson's, and the elasticity equations. *Numer. Meth. PDEs.*, 12 : 347–392, 1996.
- [3] R. E. Bank and D. J. Rose. Some error estimates for the box scheme. *SIAM J. Numer. Anal.*, 24 : 777–787, 1987.
  - [4] T. Barth and M. Ohlberger. *Finite volume methods : foundation and analysis*. Encyclopedia of computational Mechanics, volume 1, chapter 15. John Wiley & Sons, 2004.
  - [5] J. Bramble and A. Schatz. High order local accuracy by averaging in the finite element method. *Math. Comp.*, 31 : 94–111, 1997.
  - [6] Z. Cai. On the finite volume element method. *Numer. Math.*, 58 : 713–735, 1991.
  - [7] Z. Cai and J. Douglas and M. Park. Development and analysis of higher order finite volume methods over rectangles for elliptic equations. *Adv. Comput. Math.*, 19 : 3–33, 2003.
  - [8] W. Cao and Z. Zhang and Q. Zou. Superconvergence of any order finite volume schemes for 1D general elliptic equations. *J. Sci. Comput.*, 56 : 566–590, 2013.
  - [9] W. Cao and Z. Zhang and Q. Zou. Finite volume superconvergence approximation for one-dimensional singularly perturbed problems. *J. Comput. Math.*, 31 : 488–508, 2013.
  - [10] C. Chen. *Structure Theorey of Superconvergence of Finite Elements* (in Chinese). Hunan Science and Technology Press, Hunan, China, 2001.
  - [11] C. Chen and Y. Huang. *High accuracy theory of finite elements* (in Chinese). Hunan Science and Technology Press, Hunan, China, 1995.
  - [12] C. Chen and S. Hu. The highest order superconvergence for bi- $k$  degree rectangular elements at nodes- a proof of  $2k$ -conjecture. *Math. Comp.*, 82 : 1337–1355, 2013.
  - [13] J. Douglas and T. Dupont. Galerkin approximations for the two point boundary problem using continuous, piecewise polynomial spaces. *Numer. Math.*, 22 : 99–109, 1974.
  - [14] L. Chen. A new class of high order finite volume methods for second order elliptic equations. *SIAM J. Numer. Anal.*, 47 : 4021–4043, 2010.
  - [15] Z. Chen and J. Wu and Y. Xu. Higher-order finite volume methods for elliptic boundary value problems. *Adv. Comput. Math.*, 37 : 191–253, 2012.
  - [16] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. 2nd Ed., Academic Press, Boston, 1984.
  - [17] Ph. Emonot. *Methods de volumes elements finis : applications aux equations de navier-stokes et resultats de convergence*. Lyon, 1992.
  - [18] R. Ewing and T. Lin and Y. Lin. On the accuracy of the finite volume element based on piecewise linear polynomials. *SIAM J. Numer. Anal.*, 39 : 1865–1888, 2002.
  - [19] R. Eymard and T. Gallouet and R. Herbin. *Finite Volume Methods*. In : *Handbook of Numerical Analysis*, VII, 713–1020, P. G. Ciarlet and J. L. Lions Eds., North-Holland, Amsterdam, 2000.
  - [20] M. Křížek and P. Neittaanmäki. On superconvergence techniques. *Acta Appl. Math.*, 9 : 175–198, 1987.
  - [21] R. Li and Z. Chen and W. Wu. *The Generalized Difference Methods for Partial differential Equations*. Marcel Dikker, New Youk, 2000.
  - [22] J. Lv and Y. Li.  $L^2$  error estimates and superconvergence if the finite volume element methods on quadrilateral meshes. *Adv. Comput. Math.*, 37 : 393–416, 2012.
  - [23] C. Ollivier-Gooch and M. Altena. A high-order-accurate unconstructed mesh finite-volume scheme for the advection-diffusion equation. *J. Comput. Phys.*, 181 : 729–752, 2002.
  - [24] M. Plexousakis and G. Zouraris. On the construction and analysis of high order locally conservative finite volume type methods for one dimensional elliptic problems. *SIAM J. Numer. Anal.*, 42 : 1226–1260, 2004.
  - [25] A. H. Schatz and I. H. Sloan and L. B. Wahlbin. Superconvergence in finite element methods and meshes which are symmetric with respect to a point. *SIAM J. Numer. Anal.*, 33 : 505–521, 1996.
  - [26] E. Süli. Convergence of finite volume schemes for Poisson's equation on nonuniform meshes. *SIAM J. Numer. Anal.*, 28 : 1419–1430, 1991.
  - [27] V. Thomee. High order local approximation to derivatives in the finite element method. *Math. Comp.*, 31 : 652–660, 1997.
  - [28] L. B. Wahlbin. *Superconvergence in Galerkin finite element methods*. Lecture Notes in Mathematics, Vol. 1605, Springer, Berlin, 1995.
  - [29] J. Xu and Q. Zou. Analysis of linear and quadratic simplital finite volume methods for elliptic equations. *Numer. Math.*, 111 : 469–492, 2009.
  - [30] Z. Zhang and Q. Zou. Finite volume schemes of any order over rectangles. *J. Sci. Comput.*, DOI 10.1007/s10915-013-9737-5, 2013.
  - [31] J. Zhou and Q. Lin. Bi- $p$  conjecture of superconvergence-weighted norm estimates of discrete Green funtion (in Chinese). *Mathematics in Prattice and Theory*, 37 : 87–94, 2007.

- [32] Q. Zhu and Q. Lin. *Superconvergence Theory of the Finite Element Method* (in Chinese). Hunan Science and Technology Press, Hunan, China, 1989.
- [33] Q. Zou. Hierarchical error estimates for finite volume approximation solution of elliptic equations. *Appl. Numer. Math.*, 60 : 142–153, 2010.